Wang, K., Zhang, J., Shen, Y., Karkera, B. M., Croxford, A. J., & Wilcox, P. D. (2021). Defect Detection in Guided Wave Signals Using Nonlinear Autoregressive Exogenous Method. *Structural Health Monitoring: An International Journal*. https://doi.org/10.1177/14759217211018698

Title:

Defect Detection in Guided Wave Signals Using Nonlinear Autoregressive Exogenous Method

Authors and Addresses:

Kangwei Wang[1,2], Jie Zhang[2], Yi Shen[1], Benjamin Karkera[2], Anthony J. Croxford[2], Paul D. Wilcox[2]

[1]Department of Control Science and Engineering,

Harbin Institute of Technology, Harbin 150001, P. R. China

[2]Department of Mechanical Engineering, University Walk,

University of Bristol, Bristol BS8 1TR, UK

E-mails: kwwang@hit.edu.cn

Keywords:

Defect detection, Structural Health Monitoring, Guided wave, Time series modelling, ROC curve

Abstract

To perform long-term structural health monitoring (SHM), a method based on a Nonlinear Autoregressive Exogenous (NARX) network is used to learn the features present in signals acquired from a pristine structure. When a subsequent measured signal is input to the trained NARX network, the output is a prediction of the equivalent signal from a pristine structure. The residual when the pristine predicted signal is subtracted from the measured signal is used for defect detection and localization. A methodology of how to train, test and assess a NARX network for guided wave signals is introduced and applied to experimental data obtained over a period of 8 years from a sparse array of guided wave sensors deployed on a steel storage tank. A separate NARX model is trained for each sensor pair in the array using data captured in 2012. The method is first tested using data from a single pair of sensors. Defect signals are synthesized by superposing simulated responses from defects onto later experimental signals obtained from the real structure. The test results for the NARX method show better detection performance than those from the Optimal Baseline Selection (OBS) method, in terms of Receiver Operating Characteristic (ROC) curves. The detection performance of NARX method is further assessed on signals from the whole sensor array, again with simulated defect responses superposed. It is shown that good detection and localization performance can be achieved by combining the NARX residual signals from different sensor pairs. The NARX method is tested on experimental data acquired at intervals over the following 7 years as the condition of the tank naturally degrades. Indications from localized corrosion are observed. Finally, an artificial localized anomaly is added to the tank and is visible at the correct location in the image formed using the NARX method.

1 INTRODUCTION

Guided wave structural health monitoring (GWSHM) has the potential to enable complete inspection coverage of large structures with a sparse array of permanently installed transducers. GWSHM has been investigated for monitoring high-pressure vessels, oil and gas pipelines, railways, and aerospace structures [1-6]. In the generic GWSHM sparse sensor array considered here, each sensor can act as either a transmitter or receiver and an automated data acquisition system periodically records the time-domain signal from each possible transmit-receive sensor pair in the array. The received signals are

normally dominated by the direct transmission from a transmitter and subsequent reflections from benign structural features, such as edges and flanges. Because of the inherent complexity of signals resulting from these structural responses, defect detection is generally achieved by subtracting a baseline signal from the measured signal, where the baseline was acquired early in the life of the structure when it is assumed pristine. It is well known that this subtraction process is challenging due to the presence of changing environmental and operational conditions (EOCs) that alter measured guided wave signals even in the absence of any damage [7]. In particular, temperature variance is one of the dominant factors that leads to mode-dependent changes in guided wave characteristics such as wave velocity and attenuation. These effects must be compensated for to ensure baseline subtraction can achieve sufficient suppression of structural responses to reveal the presence of additional weak signals due to damage [8]. One solution is to use optimal baseline selection (OBS) [9]. In this approach, the baseline that best matches the current measurement is selected from an ensemble collected under varying EOCs for each sensor pair. A baseline may also be subjected to further manipulation (e.g. amplitude scaling, time dilation, time delay and phase adjustment) to improve the match. Such manipulations can be described as global manipulations, in that they are applied consistently to the whole baseline signal, which is then subtracted from the current measurement. Therefore, the implicit assumption in this approach is that EOCs lead to global distortions of the guided wave signal that can be described by a small number of parameters [10]. For spatially homogeneous variations in temperature, this is a reasonable assumption. The consequence is that the full range of EOCs can be assumed to be spanned by a modest-sized set of baselines and fine-tuning by global manipulation of the best-matched baseline is justified. However, the reality in most cases is that EOCs cause local rather than global distortions of signals and require a large number of parameters to be adequately described. Obtaining sufficient baselines to adequately span the multi-dimensional parameter space of EOCs is challenging (especially since the number of dimensions is not known) and global manipulations of baseline signals to improve the match are less effective. Compensation of local rather than global changes in baseline signals has been used in more recent methods of baseline signal manipulation such as the Location Specific Temperature Compensation method and dynamic time warping method [11-13].

In the current article, data-driven machine learning is applied to the problem of baseline signal subtraction. Several researchers have previously applied data-driven machine learning algorithms to different aspects of GWSHM problems. Ying *et al.* used adaptive boosting and a support vector machine (SVM) to develop a defect classifier for damage feature identification and automatic selection [14]. Liu transformed GW signals to short time Fourier transform (STFT) spectrum images and used them to train a convolutional neural network (CNN) to detect notch-type cracks in thin plate structures [15]. Hesser *et al.* used simulated impact damage to train a multi-layer perceptron to localize the impact positions in a plate [16]. Todd *et al.* employed a Bayesian neural network surrogate model with variational inference, learning latent damage features to monitor length of unexpected gaps between the lock wall and the quoin block in cargo ships [17]. Although these works present quite high detection accuracy in short-term random tests, the classification models employed cannot guarantee good performance in long-term tests with realistic variance of EOCs.

Defect detection based on classifier models has two intrinsic problems. First, they have to be designed to detect specific types of defect at specific locations and orientations (e.g. notch-type cracks [15], impact damage [16]). Training the classifiers therefore requires both pristine (defect-free) data and data containing defect responses at multiple locations and orientations. The latter is extremely costly to acquire experimentally on real structures, and modelling is also challenging since the parameter space of possible defect types, locations and orientations is very large. Secondly, both the pristine and defect data needed for training should also contain the appropriately varying EOCs if the classifier is to be used for long-term surveillance.

In the current paper, a fundamentally different machine learning approach is applied that avoids the requirement for training data from structures containing defects. Instead, a neural network is trained to predict the pristine guided wave signal from a structure given a measured guided wave signal as input. The predicted pristine signal is then subtracted from the measured signal to reveal the presence of any defect responses. The advantage of this approach is that the training can be performed entirely from measured signals recorded in the early stages of monitoring when a structure is assumed to be pristine. In this respect, the approach is similar to OBS, but instead of recording baseline signals in the pristine state and using these later, the baseline signals are used to train a neural network; instead of subtracting

the best-matched measured baseline, the pristine signal predicted by the trained neural network is subtracted. The hypothesis is that a neural network trained using pristine signals obtained under varying EOCs can learn a better representation of the effect of EOCs with local variations in baselines than the OBS approach. From this starting point, the first question is to decide the type of neural network to use and crucially how to train it. In the training stage, the input to the neural network is a measured signal from the pristine structure and the desired output is a prediction of the same, hence the training objective is to minimize the difference between the two. The neural network must be forced to make some sort of prediction of every value in the output signal to prevent it simply learning to make the output an exact copy of the input. In theory, this could be achieved by an autoencoder neural network with a sufficiently large latent parameter space to describe the EOCs, but the number of parameters is unknown *a priori*. Instead, the inherent causality in a time-domain guided wave signal is exploited: if a defect is introduced into a structure it cannot introduce any change into a measured guided wave signal before a certain instant in time associated with the propagation delay of a guided wave travelling from the transmitting sensor to the defect location and back to the receiving sensor. Therefore, a neural network (trained on pristine data) to predict future values in a time-domain signal from previous values will not be able to predict the start of the change in signal due to the presence of a defect. The prediction of future values in a time series is a nonlinear auto-regression problem and for this reason, a recurrent neural network (RNN) architecture is selected, as they can be applied to capture the dynamic behavior patterns from recursive lagged input terms (and lagged output terms if necessary) [3]. Here, a Nonlinear Autoregressive Exogenous (NARX) architecture is used, as this has been used to produce a RNN that can be efficiently trained for broadly similar applications. For example, Zhang *et al.* applied a NARX network to the analysis of acoustic emission signals to detect cracks in rail [18]. Yan *et al.* developed a damage detection approach for vibrational systems based on a NARX network. A statistical indicator based on F-test was employed with the prediction error calculated using NARX to identify abrupt changes of the prediction error variance from structural damage [19]. Baca Ruiz showed that exploiting the exogenous data input to a NARX network enabled more efficient training and accurate prediction when compared to an equivalent plain non-linear autoregressive network [20].

In the current paper, NARX processing is applied to GWSHM data obtained at various intervals over an 8-year period from a sparse array of piezoelectric sensors installed on a steel water tank. The NARX network design and tuning is described and the network is trained on 20 datasets obtained in 2012. In the current paper, the tuning of the NARX network is determined using mutual information and Cao's method [21]. The NARX performance is quantitatively assessed and compared to OBS using simulated defect responses superposed onto measured signals obtained in subsequent years. The detection performance comparison is performed using receiver operating characteristic curves (ROCs) initially for a single sensor pair in the SHM system. The ROCs for all sensor pairs in the sparse array are then computed and compared. The detection performance of the whole sensor array is demonstrated on data acquired over the following 7 years as the condition of the tank naturally deteriorates. Finally, an artificial localized anomaly is introduced into the structure and it is shown that the NARX method enables it to be detected and localized by the GWSHM sensor array.

2 METHODOLOGY OF DEFECT DETECTION USING GUIDED WAVES

2.1 Overview of defect detection process

Fig. 1(a) shows a schematic diagram of the general process of defect detection using either the OBS or NARX method. For a sparse array of sensors, the methods are applied separately to data from each pair of sensors, hence in this section the sensor indices are omitted for brevity.

For both methods, the input is a measured time-domain guided wave signal, $s(t)$ and the objective is to obtain the equivalent pristine baseline signal, $\hat{s}(t)$, such that the residual signal, $e(t) = s(t) - \hat{s}(t)$, is zero if no defect responses are present in the measured signal. Both methods rely on a library of N_{train} signals, $\mathbf{s}^{(train)} = \{s_{k_{train}}(t)\}$ where $k_{train} = 1 \dots N_{train}$, recorded under varying EOCs when the structure is assumed to be pristine. The basic principle of the OBS method is to obtain the equivalent pristine baseline signal, $\hat{s}^{(OBS)}(t)$, by selecting the signal in $\mathbf{s}^{(train)}$ that, after some manipulation, best-matches $s(t)$. This is described in more detail in Section 2.2. Conversely, in the NARX method, the data in $\mathbf{s}^{(train)}$ is used to train a NARX network to predict the equivalent pristine baseline signal, $\hat{s}(t)$, when a new signal, $s(t)$, is input.

2.2 Optimal baseline selection method

The OBS methodology used in the current paper is summarized below; further details of how the time-dilation is implemented numerically may be found in [9]. Given a measured signal, $s(t)$, the optimal baseline, $\hat{s}^{(OBS)}(t)$, is given by:

$$\hat{s}^{(OBS)}(t) = A_m s_{k_m}(\beta_m\, t), \tag{1}$$

where

$$\{A_m, k_m, \beta_m\} = \underset{\{A,k_{train},\beta\}}{\operatorname{argmin}} \left(\max_t \left| s - A s_{k_{train}}(\beta t) \right|\right).$$

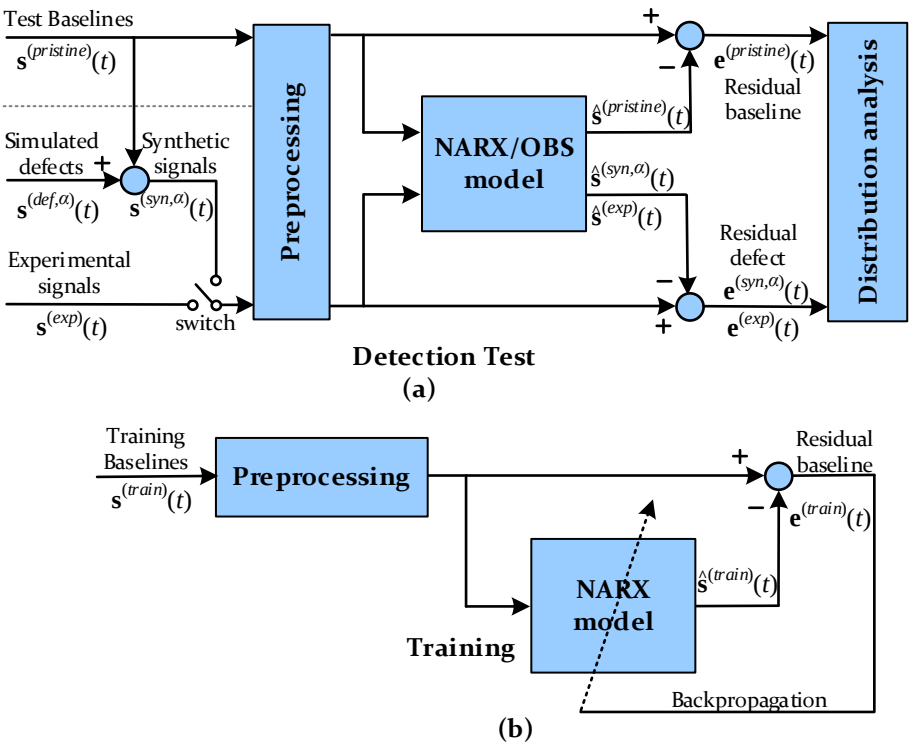
The index k_m represents the index of the best-matched baseline in $\mathbf{s}^{(train)}$ to the current measured signal. The terms A_m and β_m are respectively the optimal values of the amplitude scaling and time dilation of the best matched baseline signal to improve the match. Ideally, the residual signal from the subtraction $s - \hat{s}^{(OBS)}$ suppresses reflections from benign structural features but preserves additional signals due to scattering from defects that are not present in the pristine structure. Note that a computationally less demanding version of the OBS approach is to first identify the best-matched unaltered baseline and then optimize the time-dilation and amplitude scaling of that baseline to improve the match. In the approach used here, the selection of baseline, amplitude scaling and time-dilation are optimized together, which is guaranteed to achieved equal or superior performance.

2.3 Nonlinear Autoregressive Exogenous (NARX) method

It is more appropriate in this section to use discrete time notation denoted by square brackets where $[n] \equiv (n\Delta t)$. The objective is to obtain a neural network that approximates the function f such that $s[n+1] = f(s[n-n_h], \ldots, s[n])$, where n_h is the number of previous values considered. This is a nonlinear autoregression problem. To facilitate efficient design of a network, Takens' embedding theory is exploited [23]. This enables full use to be made of the computational ability of the NARX structure to solve the latching problem, caused by the vanishing gradient in time, determine hyperparameters, and learn long-term dependencies in the input efficiently. It has been previously shown [24] that a nonlinear autoregression network compatible with this formulation can be implemented in the Matlab

Deep Learning Toolbox using the architecture of a Nonlinear Autoregressive Network with eXogeneous inputs (NARX) with a single hidden layer. The exogeneous input to a NARX network enables the previous values of the input sequence to be passed in the appropriate form. For consistency with previous work [18, 25] this is referred to here as a NARX network, even though its overall functionality in terms of input and output does not involve any exogeneous data.

A schematic diagram of the NARX network in the training stage is shown in Fig. 1(b). The input signal to the NARX network is always normalised by its own maximum at the beginning of the NARX process and this is indicated by the pre-processing in Figs. 1(a) and (b).



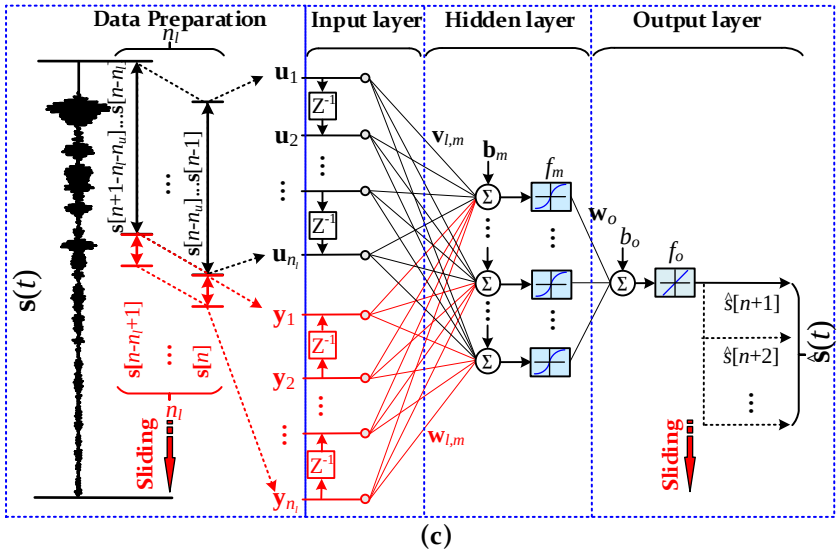


Figure. 1 The schematic diagram of; (a) a general process of defect detection, (b) NARX training and (c) the architecture of the NARX model. $\mathbf{s}^{(syn,\alpha)}$ in (a) contains synthetic datasets made by superposing different simulated defect responses, $\mathbf{s}^{(def,\alpha)}$, $\alpha = 1{,}2$ to baselines, explained in detail in section 3.3.

The NARX network architecture used is shown in Fig. 1(c). The network has three layers, an input layer, a hidden layer, and an output layer. In general terms the operation is as follows. A certain collection of n_h values from the input sequence, $\mathbf{s}$, up to and including $s[n]$ are first extracted and re-grouped to form the input layer. The information is passed through the hidden layer and the output layer to deliver the predicted value, $\hat{s}[n+1]$. The process is repeated for each value of n to build up the predicted signal, $\hat{s}$ (which is n_h elements shorter than s, because the first predicted value in $\hat{s}$ is equivalent to the $(n_h+1)^{\text{th}}$ value in s; the first n_h values in s are therefore discarded prior to evaluating the residual $e = s - \hat{s}$).

The distinguishing feature of the network that can be implemented within a NARX architecture is the way in which the input layer values are selected from $\{s[n-n_h], \ldots, s[n]\}$ to enable Takens' embedding theorem to be exploited. The input layer is connected to two tapped delay lines, $\mathbf{u}$ and $\mathbf{y}$. In the general case of predicting the series at $s[n+n_f]$ (i.e. n_f points ahead of the current point) the input, $\mathbf{y}$, contains known past values from the series and the previously predicted future values, i.e. $\mathbf{y} \equiv \{s[n-n_l+1], \ldots s[n], \hat{s}[n+1], \ldots, \hat{s}[n+n_f-1]\}$, where n_l is termed the time delay and is defined below. The second input, $\mathbf{u}$, would normally contain tapped delay lines from one or more exogenous

sources. Here, this contains n_l tapped delay lines, $\mathbf{u} = \{\mathbf{u}_l\}$, where $l = 1 \ldots n_l$, each containing n_u previous values of $\mathbf{s}$ according to:

$$\mathbf{u}_l = \{s[n + l - n_l - n_u], \ldots, s[n + l - n_l - 1]\}, \tag{2}$$

The quantity n_u is termed the embedding dimension. In the general case of multi-step ahead prediction the input layer of the network is connected to $\{\mathbf{u}, \mathbf{y}\}$. Here, only the next point is predicted (i.e. $n_f = 1$), hence the input to the network is:

$$\{\mathbf{u}, \mathbf{y}\} = \begin{Bmatrix} s[n+1-n_l-n_u], \ldots, s[n-n_l], & s[n-n_l+1], \\ \vdots & \vdots \\ s[n+l-n_l-n_u], \ldots, s[n+l-n_l-1], & s[n+l-n_l], \\ \vdots & \vdots \\ s[n-n_u], \ldots, s[n-1], & s[n] \end{Bmatrix} =$$

$$\begin{Bmatrix} s[n+1-n_l-n_u], \ldots, s[n-n_l+1], \\ \vdots \\ s[n+l-n_l-n_u], \ldots, s[n+l-n_l], \\ \vdots \\ s[n-n_u], \ldots, s[n] \end{Bmatrix}, \tag{3}$$

The output from the hidden layer can be expressed as,

$$\mathbf{h} = f_h\left(\mathbf{v}^{\mathrm{T}}\mathbf{u} + \mathbf{w}^{\mathrm{T}}\mathbf{y} + \mathbf{b}\right), \tag{4}$$

where $\mathbf{v}$ is the $(n_l n_u) \times n_m$ weighing matrix associated with the $\mathbf{u}$ input, $\mathbf{w}$ is the $n_l \times n_m$ weighting matrix associated with the $\mathbf{y}$ input, $\mathbf{b}$ are biases, n_m is the number of nodes in the hidden layer, f_h is the activation function used in the hidden layer nodes, and superscript T denotes transpose. The elements in $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are learned when the network is trained. In the output layer, the output is given by

$$\hat{s}[n+1] = f_o\left(\mathbf{w}_o^{\mathrm{T}}\mathbf{h} + b_o\right), \tag{5}$$

where $\mathbf{w}_o$ is the $n_m \times 1$ weighting matrix, b_o is the bias, and f_o is the activation function in the output layer. The total number of parameters that must be determined in training is $[n_l(n_u + 1) + 2]n_m + 1$. The activation functions are tangent sigmoid functions for the hidden layer nodes and linear for the output layer.

The NARX network is trained to minimise the residual mean squared error (MSE) between the input training signals $\mathbf{s}^{(train)}$ and the predicted output, $\hat{\mathbf{s}}^{(train)}$, by optimized parameters, $\boldsymbol{\theta} = [\mathbf{v}, \mathbf{w}, \mathbf{w}_o, \mathbf{b}, b_o]$:

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{k_{train},n} \left| s_{k_{train}}[n] - \hat{s}_{k_{train}}[n] \right|^2. \tag{6}$$

After obtaining $\boldsymbol{\theta}$, the detection performance of the NARX has to be tested using,

$$\mathbf{e}^{(test)} = \left| \mathbf{s}^{(test)} - \hat{\mathbf{s}}^{(test)} \right|, \tag{7}$$

where $|.|$ denotes an envelope function that returns all the absolute values of the vector elements. Note that, $\mathbf{s}^{(test)}$ could be any signal from the structure at any conditions (with or without a defect).

2.4 Determination of NARX structure

In common with the primary challenge for all neural network techniques, the selection of the hyperparameters of the NARX structure is crucial and requires investigation. As described, the length of considered previous values in NARX, $n_h = n_l + n_u$, is invariant and it is thus crucial to choose an appropriate length. Here two critical parameters are involved, the time delay n_l and embedding dimension n_u. Intuitively, the relative sizes of n_l and n_u determines whether the short-term or long-term information has more contribution to recreating the dynamic behaviour of the signal. With an appropriate selection of these two parameters, the dynamic characteristics of the time series should be captured in the embedding space. In order to optimise this parameter selection, Gao *et al.* used a least squares support vector machine (LSSVM) for parameter fine-tuning and optimized their fault detection model using an adaptive genetic algorithm [26]. Perez-Ramirez proposed a mutual information index to ascertain the optimum number of neurons in the layers of the NARX [27]. In both cases their selections were specific to their applications and are not generally applicable. Instead of tackling the problem as an optimisation process, this paper determines the parameters using the delay embedding theorem and ascertains the optima based on the physical characteristics of the original signals. In the literature, the estimation of n_l based on mutual information analysis has proven to be most accurate for Takens' theorem and is the method employed here.

The optimum time delay, n_l, is calculated by the first minimum value of average mutual information $I(n_l')$ between amplitude distributions of the signal $s[n]$ and its lagged signal $s[n+n_l']$. $I(n_l')$ is defined as:

$$I(n_l') = \sum_{p,q} P_{pq}\{s[n], s[n+n_l']\} \cdot \log\left\{\frac{P_{pq}\{s[n],s[n+n_l']\}}{P_p\{s[n]\}P_q\{s[n+n_l']\}}\right\}, \quad (8)$$

where $P_p\{s[n]\}$, $P_q\{s[n+n_l']\}$ are the probabilities of finding a time series value in the p^{th} interval of the $s[n]$ amplitude distribution and the q^{th} interval of the $s[n+n_l']$ amplitude distribution respectively. $P_{pq}\{s[n], s[n+n_l']\}$ is the joint probability that an observation falls in both the corresponding intervals. In such a way, an appropriate time delay, n_l, is obtained, which guarantees a low correlation between the original signal and the signal lagged by n_l.

The embedding dimension, n_u, is estimated based on the optimal time delay using Cao's method [21]. This approach has been proposed to determine the minimum embedding dimension from a scalar time series. Consider a set of time delayed vectors,

$$\boldsymbol{\gamma}(n, n_l, n_u') = \{s[n], s[n+n_l], \cdots, s[n+(n_u'-1)n_l]\}, \quad (9)$$

where $n = 1 \ldots n_s - (n_u' - 1)$ and n_s is the number of points in s. Then the optimal embedding dimension, n_u, is determined by the evolution of the following objective,

$$E(n_u') = \frac{1}{n_s - n_l n_u'} \sum_{n_0=1}^{n_s - n_l n_u'} \frac{\|\boldsymbol{\gamma}(n,n_l,n_u'+1) - \boldsymbol{\gamma}(n_1,n_l,n_u'+1)\|}{\|\boldsymbol{\gamma}(n,n_l,n_u') - \boldsymbol{\gamma}(n_1,n_l,n_u')\|}, \quad (10)$$

where $\|\cdot\|$ returns the Euclidean norm of a vector, n_1 is an integer such that $\boldsymbol{\gamma}(n_1, n_l, n_u')$ is the nearest neighbour of $\boldsymbol{\gamma}(n_0, n_l, n_u')$ [22]. A nearest neighbour is a lagged vector close to the original vector, which presents similar topological (dynamic) properties in the delay embedding space. The embedding dimension is determined when $E(n_u')$ stops changing for any n_u'. In practice $E(n_u')$ is considered to have converged when the change is less than 1×10^{-3} [28]. This indicates the minimum length to preserve the sequence information and recreate the nonlinear dynamics of the original signal.

3 EXPERIMENTAL MEASUREMENT SYSTEM AND DATA

3.1 Experimental measurement system

Fig. 2(a) shows a photo of a water tank with a sparse array of 8 permanently attached guided wave sensors. The tank is around 2 m diameter and fabricated from 5 mm thick, semi-cylindrical steel panels that are bolted together along flanges. An unknown sealant is used between the flanges. Each sensor comprises a piezoelectric disk (NCE51, ∅20 mm × 1 mm thick, wraparound electrodes, Noliac A/S, Denmark) that is bonded to the outer surface of the tank using cyanoacrylate adhesive. The conductors from a coaxial cable are soldered to the electrodes on the piezoelectric disk and the disk is encapsulated in a silicone sealant to provide environmental protection. The disks are designed to operate predominantly in a radial mode and have high sensitivity to the S_0 guided wave mode at around 250 kHz in a 5 mm thick steel plate. At this frequency-thickness product, the S_0 mode has a relatively small out-of-plane displacement component and the dispersion is minimal [29]. In addition, it is the fastest fundamental mode; this makes the identification of the directly-transmitted signal between a pair of sensors and estimation of group velocity straightforward.

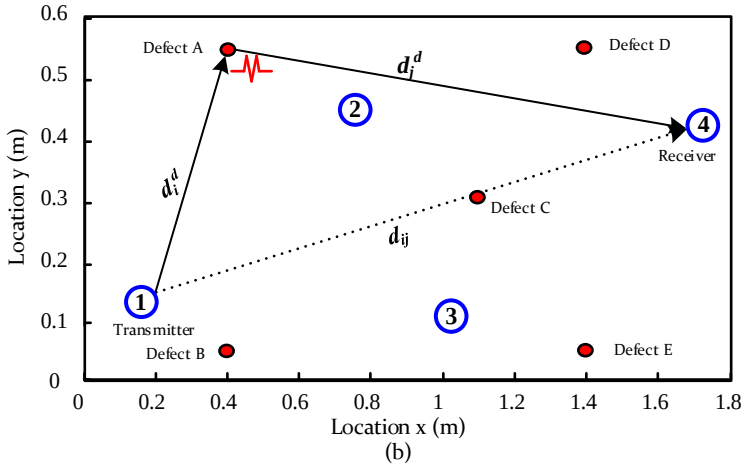
The wires from each sensor are connected in parallel to a pair of single-pole, 10-way RF relays which act as multiplexors. The pole of one multiplexor is connected to an input channel of an integrated digital oscilloscope and signal-generator device (Handyscope HS3, TiePie Engineering, Netherlands); the pole of the other multiplexor is connected to its signal-generator output. The multiplexors and oscilloscope / signal-generator device are all controlled by a PC. Periodically, the PC activates a new measurement cycle, denoted by index k. The multiplexors connect the i^{th} sensor to the signal-generator to act as transmitter and the j^{th} sensor to the oscilloscope to act as receiver. Because the output of the signal-generator is limited to 12 volts a chirp excitation function is used (1000 cycles with frequency content equivalent to a 5-cycle Hanning-windowed toneburst with centre frequency of 250 kHz) to increase the energy transmitted in the excitation signal [30]. The signal detected on the receiver is de-chirped and a single dataset is recorded into the complete set of time domain responses, $s_{ij,k}(t)$, recorded from each pair of sensors in the array during the k^{th} measurement cycle. The multiplexors then switch to the next sensor pair and the process is repeated. In one complete measurement cycle the responses from all

possible transmit-receiver pairs are acquired and stored. In the current paper to minimise the data to be displayed, only data from 4 of the 8 sensors is used, hence $i, j = 1 \ldots 4$ and 12 responses per measurement cycle are captured.

The tank is in a sheltered exterior location but has been subject to natural environmental conditions over many years. Since its installation in 2012 the GWSHM system has performed over 75000 measurements and a subset of 2400 measurements distributed through the entire acquisition period is selected in this research. When operating correctly, measurements are performed every day, but for various reasons (e.g. building work, PC failure) there are many gaps in the recorded data. Initially the experiment was stopped in July 2016 but was restarted for the purposes of acquiring new data for this paper in July 2020. Where it is necessary to refer to collections of datasets, these are indicated by superscripts $\mathbf{s}^{(X)} = \{s_{ij,k_X}\}$ where X is the name of the collection, $k_X = 1 \ldots . N_X$, and N_X is the number of datasets in X. The following collections are defined according to the calendar year in which they were acquired: $\mathbf{s}^{(2012)}$ $\mathbf{s}^{(2013)}$, $\mathbf{s}^{(2014)}$, $\mathbf{s}^{(2015)}$, $\mathbf{s}^{(2016)}$, $\mathbf{s}^{(2020a)}$ and $\mathbf{s}^{(2020b)}$. The two collections for 2020, $\mathbf{s}^{(2020a)}$ and $\mathbf{s}^{(2020b)}$, relate to measurements performed before and after the addition of a localised physical anomaly on the tank structure. Note that the tank remained empty throughout the GWSHM monitoring period.

The locations of the 4 sensors used in the current paper and the simulated defects are depicted in Fig. 2(b). The length of the recorded time traces is equivalent to a propagation path of 3.5 m, hence the region traversed by the guided waves recorded by each sensor pair covers the entire panel.

(a)



(b)

Figure. 2 (a) A photograph of the tank showing attached sensors and (b) schematic diagram illustrating the geometry of simulated defects and the sensors used in the current paper.

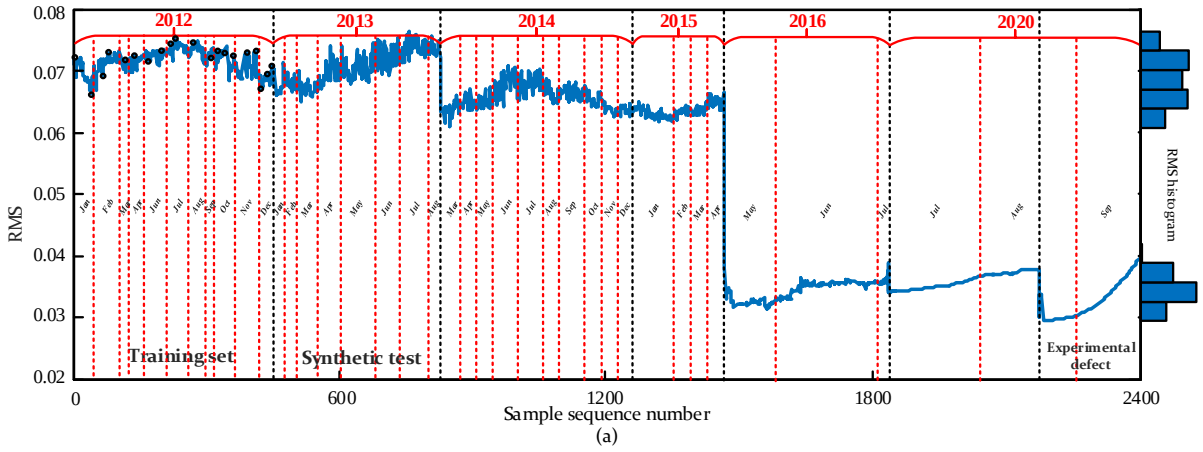
3.2 Data used for training GWSHM system

The water tank has been monitored by GWSHM since 2012. In order to demonstrate a long-time surveillance scenario, the data obtained in 2012 is used for training the GWSHM (both using conventional OBS and with NARX) system and the data from 2013 to 2020 is used for testing its defect detection performance.

Fig. 3 (a) shows the root mean square (RMS) amplitude of the acquired signals transmitted from Sensor 1 and received on Sensor 4 in chronological order and Fig. 3 (b) shows the measured group velocity for each signal. The group velocity curve is estimated based on the fixed sensor distances and time of first arrival in the baseline signals' envelope. The velocity distribution is concentrated in a relatively narrow range of 4943±6.7 m/s (about ±0.14% fluctuation). To a first approximation, the sensitivity of the S_0 group velocity to temperature will be the same as that of a bulk longitudinal wave. Based on textbook values for the sensitivity of the Young's Modulus of steel to temperature [31], this can be estimated to be ~1.3 $ms^{-1}/°C$. This is consistent with the measured variation in S_0 group velocity and the average

annual variation of temperature in Bristol, which is around 14°C. In 2012, 456 measured datasets are used over the entire year. From these, a number of datasets, N_{train}, evenly-distributed through the year were chosen as training signals (whose optimum is investigated later), $\mathbf{s}^{(train)} = \{s_{ij,k_{train}}\}$ where $i,j = 1 \dots 4$ and $k_{train} = 1 \dots N_{train}$. For the case of $(i,j) = (1,4)$, and N_{train} = 20, the signals in $\mathbf{s}^{(train)}$ are marked by circles in Fig. 3. In this data, seasonal trends in temperature can be observed as well as a general long-term reduction in amplitude, which due to deterioration of the tank and/or the sensors. It is thus not clear that a GWSHM scheme trained on data from 2012 would be sufficient to analyse later data.

For the purposes of GWSHM, these training sets are assumed to come from the tank in pristine condition under varying EOCs. For OBS, $\mathbf{s}^{(train)}$ forms the baseline dataset. For NARX, $\boldsymbol{s}^{(train)}$ is first used to investigate the basic dynamic properties of the guided wave time-traces in order to obtain the adjustable parameters for the NARX network. The same data is then used to train separate NARX networks for each of the 12 sensor pairs.



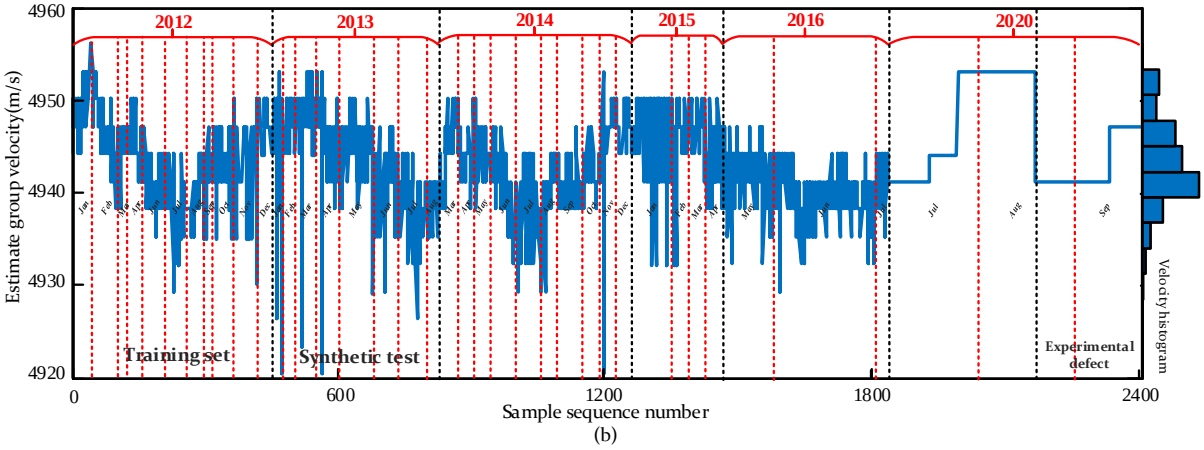


Figure. 3 (a) The root mean square (RMS) amplitude of the signals transmitted from Sensor 1 and received on Sensor 4 and (b) the estimate group velocity. Signals from the datasets used for training the GWSHM system are indicated by black circles and test sets used for the synthetic test is also illustrated.

3.3 Data used for testing GWSHM system

After the acquisition of training data, the detection performance of the NARX and OBS methods is assessed. The experimentally measured data acquired between 2012 and 2020 (indicated in Fig. 3) was subject to naturally varying EOCs and is used for this analysis. The data is split into the following three classes:

1. $\mathbf{s}^{(2013)} = \{s_{k_{test}}\}$ where $k_{test} = 1 \ldots 376$: datasets measured experimentally in 2013. These datasets are assumed to still come from a near-pristine structure and provide the basis of the synthetic defect signals used to assess detection performance.
2. $\mathbf{s}^{(2014)}$ (427 datasets), $\mathbf{s}^{(2015)}$ (204 datasets), $\mathbf{s}^{(2016)}$ (350 datasets), $\mathbf{s}^{(2020a)}$ (343 datasets prior to addition of localised anomaly) are used to assess the network's performance when subject to long-term environmental degradation.
3. $\mathbf{s}^{(2020b)}$ (241 datasets) after a localised physical anomaly (a $\varnothing 16 \times 20$ mm cylindrical steel mass bonded on the tank surface) was introduced at (1.1 m, 0.3 m), which is the same location as the modelled defect C shown in Fig. 2(b). These datasets are used for verifying the ability of NARX to detect and localise a real anomaly in the structure.

3.3.1 Synthetic defect data

To carry out a systematic quantification of the performance, data with known defects is required. To do this synthetic defect data, $\mathbf{s}^{(syn,\alpha)}$ (where $\alpha = 1$ or 2 depending on the method of defect amplitude scaling), is created by combining experimentally-measured signals in $\mathbf{s}^{(2013)}$ with simulated responses from defects, $\mathbf{s}^{(def,\alpha)}$. In addition, the baselines in the synthetic test are randomly chosen from the experimentally-measured signals. To ensure uniform test set sizes for each year, the size of the synthetic test set is selected as 204, as this is the minimum size of dataset in any year (this was the size of the dataset acquired in 2015 which was small due to a significant outage due to a lab refurbishment). The simulated defect response is assumed to be a delayed and scaled copy of the transmitted signal used in the experimental measurements after de-chirping, i.e. a 5-cycle Hanning windowed toneburst with a center frequency of 250 kHz. The defect response is introduced only at the point in a signal when a direct reflection from the defect occurs, hence secondary effects, such as shadowing of reflections from structural features by a defect, are ignored. Thus, if $s^{(ref)}(t)$ represents the shape of the effective transmitted signal with unit amplitude and zero delay, the defect signal that is combined with an experimental signal is:

$$s^{(def,\alpha)}(t) = A_{ij}\sqrt{d_{ij}}a_\alpha s^{(ref)}(t - t_d)\,, \tag{11}$$

where $A_{ij} = s^{(2013)}(t_{ij})$ is the amplitude of the directly transmitted signal that arrives at time $t_{ij} = d_{ij}/c$, d_{ij} is the separation between sensors, t_d is the arrival time of the defect response, and a_α is a scale factor controlling defect amplitude that depends on the nature of the defect and its physical location relative to the sensors. The defect is assumed to behave as an omni-directional scatterer with scattering amplitude controlled by a single coefficient $10^{\beta/20}$ where the parameter β has units of decibels (dB), which describes the ratio of the amplitude of the scattered wave at unit distance from the defect to that of an incident plane wave. The position of the defect relative to the sensors determines the effect of beam spread (the reduction in amplitude of the defect response due to divergence of the

transmitted waves and divergence of the waves scattered from the defect). The beam spread function for guided waves transmitted by the i^{th} sensor to a defect and received by the j^{th} sensor:

$$B_{ij} = \frac{1}{\sqrt{d_i^d d_j^d}} = \frac{1}{d^d \sqrt{\frac{d_i^d}{d_d}\left(1-\frac{d_i^d}{d_d}\right)}}, \tag{12}$$

where d_i^d and d_j^d are the distances between the defect and the i^{th} and j^{th} sensors respectively, and $d^d = d_i^d + d_j^d$, as shown in Figure 2(b). The value of α indicates how the beam spread is accounted for.

In the first case ($\alpha = 1$), an individual time-trace is considered and it is specified that a simulated defect signal arrives at time $t^d = d^d/c$. The total propagation distance, d^d, is therefore known, but d_i^d and d_j^d are not. The only known information about defect position is that the defect must lie on an ellipse with the transmitter and receiver elements at the foci. It is conservatively assumed that the defect position on the ellipse is that associated with the minimum defect response amplitude due to beam spread. This occurs when $d_i^d = d_j^d = d^d/2$, and leads to the minimum $B_{ij} = 2/d^d = 2/(ct^d)$. In this case

$$a_1 = \frac{2}{ct^d} 10^{\beta/20}. \tag{13}$$

In the second case ($\alpha = 2$), the location of a defect in the structure is specified so d_i^d and d_j^d are known, hence:

$$a_2 = \sqrt{\frac{1}{d_i^d d_j^d}} 10^{\beta/20}. \tag{14}$$

3.4 NARX training

Based on the theories described in section 2.4, the optimal embedding dimension and time delay for the NARX networks are determined to be $n_u = 20$ and $n_l = 4$. The number of neurons, n_m, in the hidden layer should be sufficiently large to enable a function of the required complexity to be described. The empirical estimation for the number of neurons in a hidden layer used here is $n_m \geq \sqrt{(n_u + n_y)n_l + n_{ol}} + 10$ [32, 33], where $(n_u + n_y)n_l = 84$ is the number of neurons in the input layer and n_{ol} is the number of neurons in the output layer and in this case is 1. Thus $n_m = 20$ neurons

are used in the hidden layer. This leads to 1721 network parameters that must be optimized in the NARX training process.

In addition to the hyper-parameters, there are two modes used for the training of NARX: parallel mode and series-parallel mode. In series-parallel (SP) mode, the target output is available and used as the regressive vector $\boldsymbol{y}$. It is more stable than using the predicted output as a $\boldsymbol{y}$ vector. This mode is adopted here and trained as a one-step ahead predictor to alleviate error accumulation in time [24]. The Levenberg-Marquardt backpropagation algorithm is employed to train the NARX network as it is computationally efficient when the network is relatively small. Based on a previous study [34], a sufficiently small learning rate is beneficial for avoiding oscillation and ensuring a convergent training process. Accordingly, the learning rate in all trials is selected as 1×10^{-3}. An early-stopping strategy is employed in the training process of NARX to avoid overtraining due to local optima in the network parameters. In the early-stopping strategy, the training performance is monitored with the validation loss, which provides an estimate of the generalization capacity for the network. The training process is stopped when the validation performance grows continuously worse in 10 epochs. In this case, overtraining is assumed to have happened and the network parameters with the lowest validation loss before overtraining are used for tests. The overall dataset available for training is partitioned with 10% randomly selected and reserved for validation while the remaining 90% is used for the training process itself. The validation data is kept completely separate from the training process and not used in the weights updating process in NARX training.

In order to decide how many training samples should be used, a NARX network for the sensor pair $(i, j) = (1,4)$ is first tested using varying numbers of training datasets drawn from $\mathbf{s}^{(2012)}$. The test performance, expressed as mean square error between the NARX input size and output, as a function of number of training sets is shown in Fig. 4. Here the training is carried out with the same network structure while increasing the number of randomly selected samples. Error bars at each point are calculated from 10 realisations of the same training set size and the one with lowest MSE is selected in the following tests. It can be seen that the test performance plateaus when the number of training datasets is greater than 20 and hence N_{train}=20 is used to train the NARX networks for all sensor pairs.

It should be noted that although the optimal training set encompasses only 20 signals, these signals are re-grouped into local regions for input into the NARX network, as shown in Fig. 1 (c). Therefore, one time-trace with 4000 points will be partitioned into 3976 samples with one-step ahead targets. Thus the 20-signal training set represents over 79500 samples for training the NARX network.

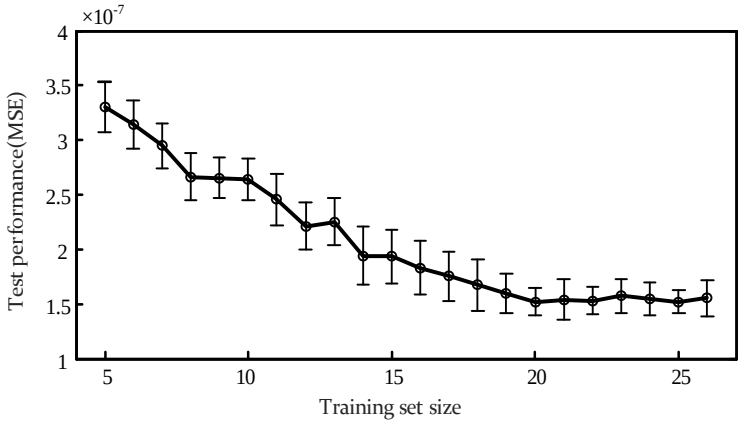


Figure. 4 Test performance of the NARX models based on different training set sizes for the signals from the sensor pair $(i, j) = (1,4)$.

3.5 Results using synthetic defect data

3.5.1 Example of NARX and OBS methods for a single sensor pair

As an example, the detection performance of both NARX and OBS is first demonstrated on synthetic signals generated using equation 13, i.e. $\alpha = 1$, for sensor pair $(i, j) = (1,4)$.

A typical test signal, $s^{(pristine)}(t)$ from $\mathbf{s}^{(2013)}$, and a simulated defect signal, $s^{(def,1)}(t)$, for a defect with an amplitude of β = -20 dB arriving at t_d = 0.506 ms are shown in Fig. 5(a). A closeup of the region around t_d is shown in Fig. 5(b), which also shows the synthetic signal, $s^{(syn,1)} = s^{(pristine)} + s^{(def,1)}$. Figs. 5 (c)-(d) show the residual signals from the NARX method for $s^{(pristine)}$ and $s^{(syn,1)}$ respectively, while Figs. 5 (e)-(f) show the residual signals from the OBS method. In this case OBS is initially applied then a baseline signal stretch to improve performance as described in Croxford et al [9]. Note that the signals in Figs. 5(c)-(f) are each normalized to the RMS amplitude of the residual signal from the defect-free case, $e^{(pristine)}$. Comparing Figs. 5(d) and 5(f), it can be seen that the NARX method shows a clear indication at the expected arrival time of the defect with a peak amplitude that is almost four times greater than the RMS level. In the OBS result, the defect response cannot be resolved

above the RMS level. While this is an encouraging result, the performance must be systematically compared.

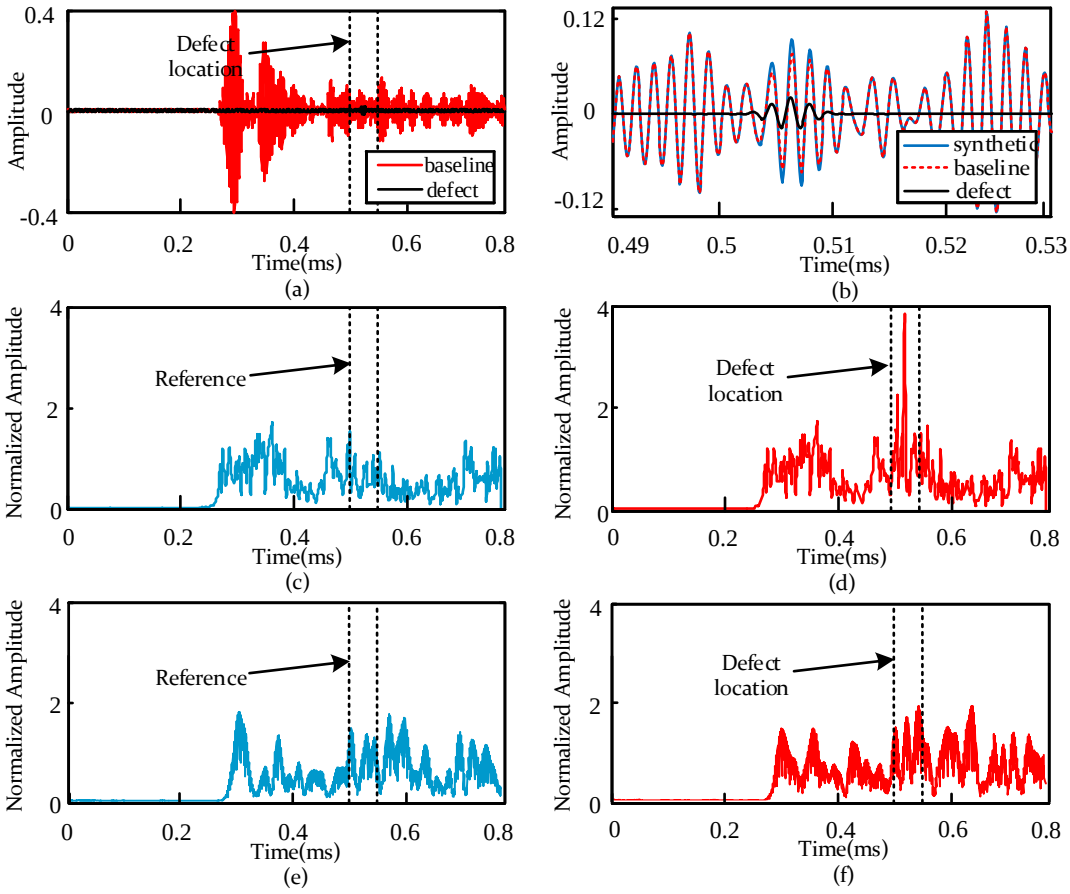


Figure. 5 Example signals: (a) an original defect-free test signal, $s^{(pristine)}$, and a simulated defect signal, $s^{(def,1)}$, (b) a closeup of (a) in the vicinity of the defect signal also showing the synthesized signal, $s^{(syn,1)} = s^{(pristine)} + s^{(def,1)}$, (c) a residual signal from a defect-free signal processed using NARX method, $e^{(pristine)}$, (d) a residual signal from a synthesized signal processed using NARX method, $e^{(syn,1)}$, (e) a residual signal from a defect-free signal processed using OBS method and (f) a residual signal from a synthesized signal processed using OBS method.

3.5.2 Assessment of NARX method performance for single sensor pair

To systematically compare the detection performance of the OBS and NARX methods for a given sensor pair, receiver operating characteristics (ROC) curves are generated and analysed using the population of 204 test signals, $\mathbf{s}^{(2013)}$. In addition, a simple regressive architecture composed of a plain neural network without the NARX structure, is also investigated here as a second comparison. This models the response of $s[n+1] = f(s[n-n_h], \ldots, s[n])$ without exogenous input and all that

implies in Section 2. The comparative network is fed with same input length of n_h, setting is identical with NARX and it is structurally referred as a Nonlinear Auto-Regressor (NAR). This is done to determine where any benefits in performance come from, the use of machine learning techniques or the specific NARX structure.

The ROC curve provides a graphical tool for visualising the performance of a binary classifier or, in the case considered here, an anomaly detector as a function of the discrimination threshold. The probability of detection (POD) for a given synthesized defect (i.e. for a simulated defect signal with a given arrival time, t_d, and amplitude, β) and the probability of false-alarm (PFA) at each threshold are computed and plotted to form an ROC curve. First each of the 204 defect-free test signals in $\mathbf{s}^{(2013)}$ is passed through the NARX and OBS models to obtain the corresponding defect-free residual signals, $\mathbf{e}^{(pristine)}$. To compute the PFA, the detection threshold is increased from zero and the fraction of signals in $\mathbf{e}^{(pristine)}$ that exceed each threshold level (i.e. the PFA for each threshold level) within a time window of 0.05 milliseconds is calculated until the threshold level is such that no residual signals exceed it. Then the same simulated defect signal, $s^{(def,1)}$, is added to each of the 204 test signals to form the set $\mathbf{s}^{(syn,1)} = \left\{s_k^{(2013)} + s^{(def,1)}\right\}$. These signals are again passed through the NARX and OBS models. The threshold is increased from zero and the fraction of residual signals in $\mathbf{e}^{(syn,1)}$ that exceed each threshold level (i.e. the POD for each threshold level) within a time window of 0.05 milliseconds calculated for all levels.

In all cases, ROC curves follow a trajectory from the origin at (0, 0), where the threshold is larger than any defect or defect-free signals, to the top right corner of the graph at (1,1), where the threshold is lower than any defect or defect-free signal. The ROC curve for a perfect detector will also pass through the top left corner of the graph at (0,1), indicating that a threshold can be found that enables a POD of 1 and a PFA of 0 to be achieved. However, when the amplitude of defect signals is reduced the amplitude distributions in the defect and defect-free cases overlap and one or both of the POD and PFA must be compromised. When comparing two different methods, the method that has a higher POD for a given PFA or a lower PFA for a given POD is the superior method. If the ROC curves of the methods cross each other, then the best method may depend on which criterion is prioritised. However, if the

ROC curves do not cross, the best method against any criterion is that with the largest Area Under the Curve (AUC), which provides a single metric for comparing methods. A perfect detector has AUC of 1 and a poor detector that fails to provide any discrimination (beyond a random guess) has an AUC close to 0.5.

Fig. 6(a) shows examples of the Probability Density Functions (PDFs) of the peak amplitudes within a time window of 0.05 milliseconds of the defect position extracted from $\mathbf{e}^{(pristine)}$ and $\mathbf{e}^{(syn,2)}$ from sensor pair $(i,j) = (1,4)$. The time window used is indicated by the vertical dashed lines in Figs. 5(d-f). The PDFs are shown for various defect amplitudes and the resulting ROC curves are shown in Fig. 6(b). It can be seen that the NARX method allows almost perfect defection of defects in this instance when the defect amplitude is -25 dB or greater. For the -30 dB defect, the PDFs for the defect and defect-free cases overlap significantly and consequently the ROC curve does not pass close to (0,1) and represents a significantly worse detector.

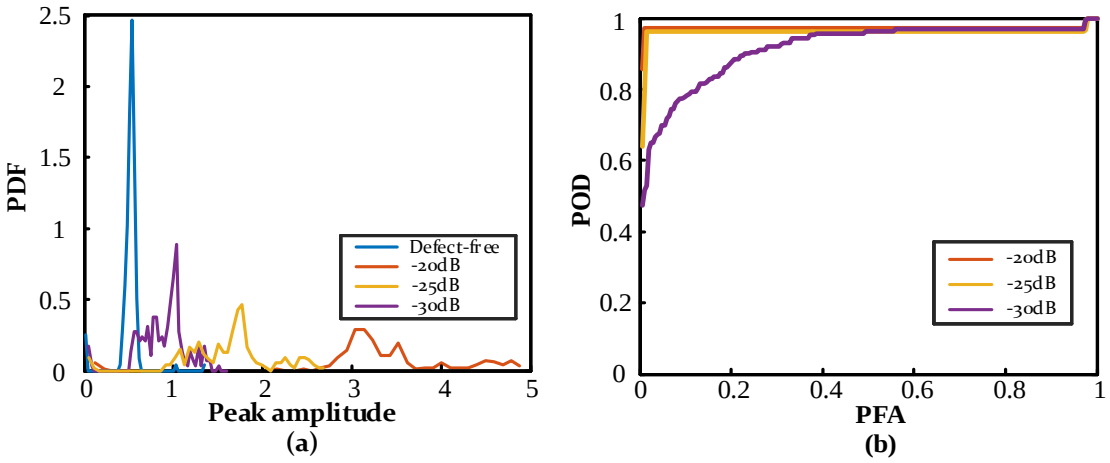


Figure. 6 (a) The PDFs of the peak amplitudes extracted from NARX $\mathbf{e}^{(pristine)}$ and $\mathbf{e}^{(syn,2)}$ at t^d = 0.506 ms from a sensor pair of $(i,j) = (1,4)$ and (b) the ROC curves generated from (a). Note that, the corresponding AUCs of -20dB, -25dB and -30dB curves in (b) are 0.9666, 0.9623 and 0.9373, respectively.

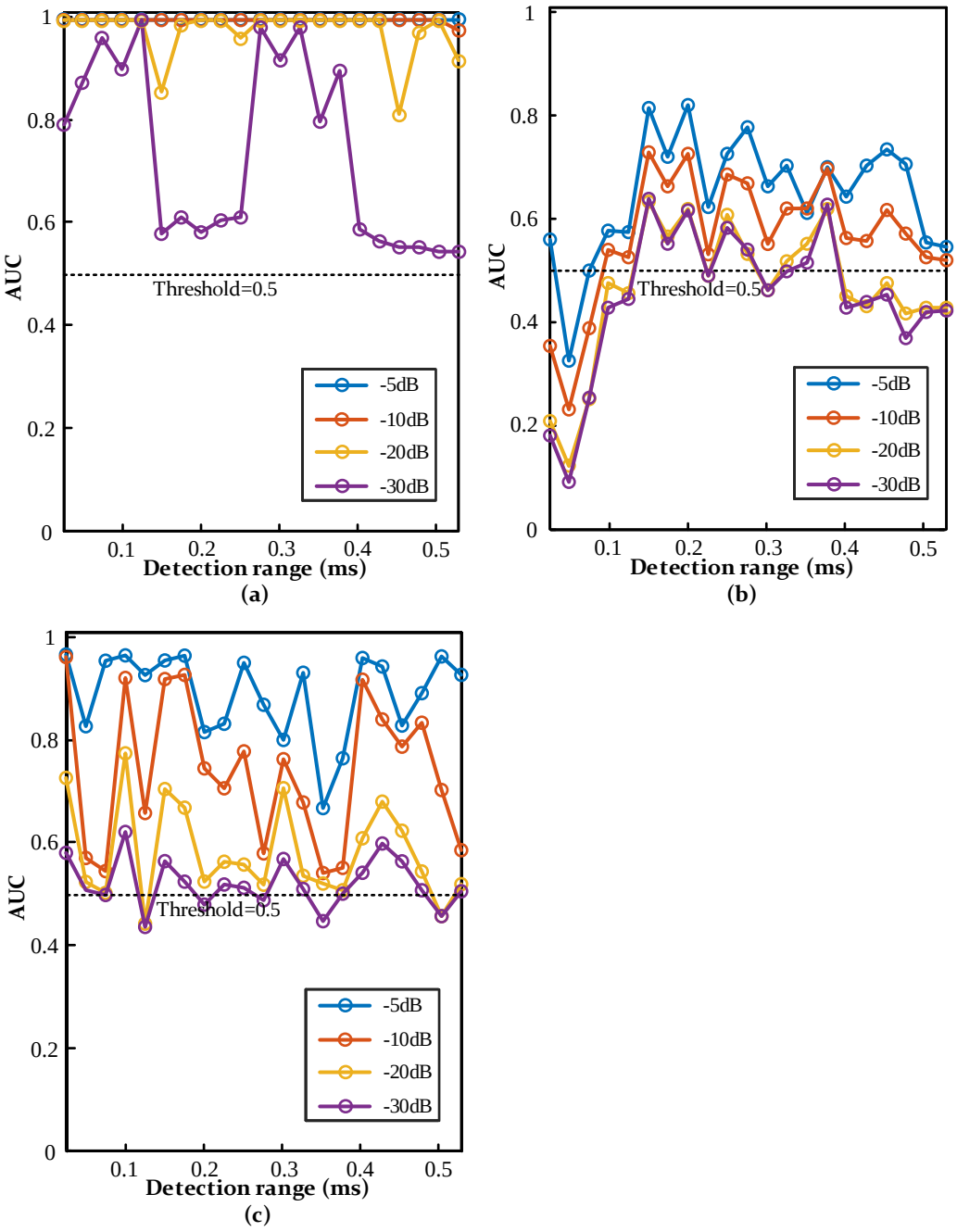


Figure. 7 The AUC as a function of detection range for $\mathbf{s}^{(syn,2)}$ processed by; (a) the NARX, (b) the OBS and (c) the NAR. Note that $\mathbf{s}^{(syn,2)}$ is the superposition of $\mathbf{s}^{(2013)}$and $\mathbf{s}^{(def,2)}$ with defect severities of β = -5 dB, -10 dB, -20 dB and -30 dB.

The detection performance of the OBS, the NAR and the NARX methods for the same sensor pair, $(i,j) = (1,4)$, is further assessed for defects at various arrival times, t_d, and amplitudes of β = -5 dB, -10 dB, -20 dB and -30 dB. The results are shown in Figs. 7. Note that the lateral axes in Figs. 7 are plotted as detection range, defined as defect arrival time after the first arrival signal time equal to $t^d - d_{ij}/c$. Comparing Figs. 7(a-b), it can be seen that:

1) when $\beta \geq$ -20dB, the OBS method performs well enough to extract the defects at various locations with an average AUC (for defects of $\beta =$ -5 dB, -10 dB and -20 dB) of 0.5679 and a standard deviation of 0.1269. The NARX method provides a near perfect detectability with an average AUC of 0.994 and a standard deviation of 0.0049.

2) When the defect amplitude decreases to a certain low level (β = -30dB), the AUC values from the OBS and NAR method are distributed around 0.5, indicating minimal detection capability. In addition, at low defect amplitudes it can be observed that the detection performance of the NARX and OBS methods have approximately opposed tendencies as a function of detection range. A possible hypothesis is that NARX performs better when the baseline has high amplitude, while OBS performs better when the baseline has low amplitude. However, the NARX method shows significantly higher AUC values than both NAR and OBS in all levels. This suggests that the excellent performance is a function of the specific structural traits of NARX, not just machine learning techniques. A reasonable architecture can enhance the efficiency of network considerably.

3.5.3 Assessment of NARX method performance for multiple sensor pairs

Due to the reciprocity principle [35,36], the signals from sensor pair (i, j) are the same as those from pair (j, i). Here the signals from 6 sensor pairs for the case of $j > i$ were used to assess the detection performance of the NARX method when applied to a network of sensors. The signals used are $\mathbf{s}^{(2013)}$ comprising 204 datasets measured in 2013 and $\mathbf{s}^{(syn,1)}$ which combines $\mathbf{s}^{(2013)}$ with the response from the modelled defects A-E in Fig. 2(b), with β ranging from -5 dB to -50 dB. Fig. 8 shows the resultant AUC for 5 defects as a function of sensor pair and β.

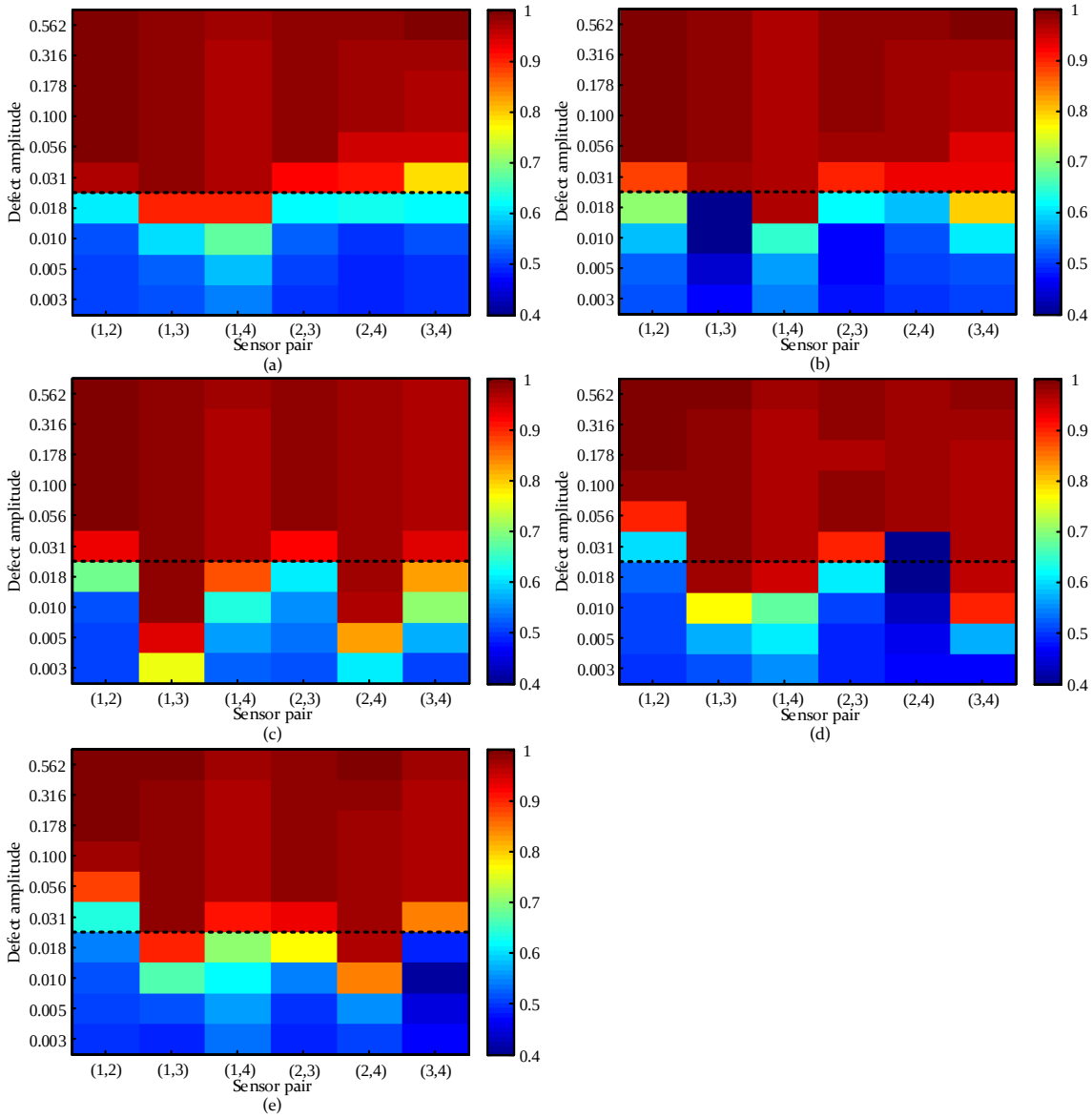


Figure. 8 Detection performance (AUC) for (a) defect A, (b) defect B, (c) defect C, (d) defect D and (e) defect E as a function of sensor pair and defect response amplitude (corresponds to the severity range from -5 dB to -50 dB in 5dB increments). Horizontal axis indicates the corresponding indices of sensor pairs, (i, j). The dashed line indicates a threshold where the detection performance for the majority of sensor pairs deteriorates rapidly.

3.6 Results from NARX method using purely experimental data

In order to further verify the practicability of the proposed method, the NARX method (using the networks trained on 2012 data) is demonstrated on experimentally measured datasets from subsequent years as the sensor array ages and the condition of the tank degrades naturally. In August 2020, a localised anomaly (a bonded mass) was added to the tank and the NARX method applied (without retraining) to the subsequent data, $\mathbf{s}^{(2020b)}$.

To give an intuitive illustration of detection results from the real defect, Figure 9(a,c) shows the residual signal after the NARX process is applied to signals from $\mathbf{s}^{(2013)}$ and $\mathbf{s}^{(2020a)}$ respectively. Figure 9(b) shows the residual signal from a $\mathbf{s}^{(syn,1)}$ composed from $\mathbf{s}^{(2013)}$ and a synthetic signal from a defect at (1.1 m, 0.3 m) with β = -30dB. Figure 9(d) shows the residual signal from a signal in $\mathbf{s}^{(2020b)}$ after the localised anomaly was added to the tank. Comparing Figures 9(b) and 9(d), it can be seen that, although visible, the response of the localised anomaly is very small. The magnitude of this response will be further explored later.

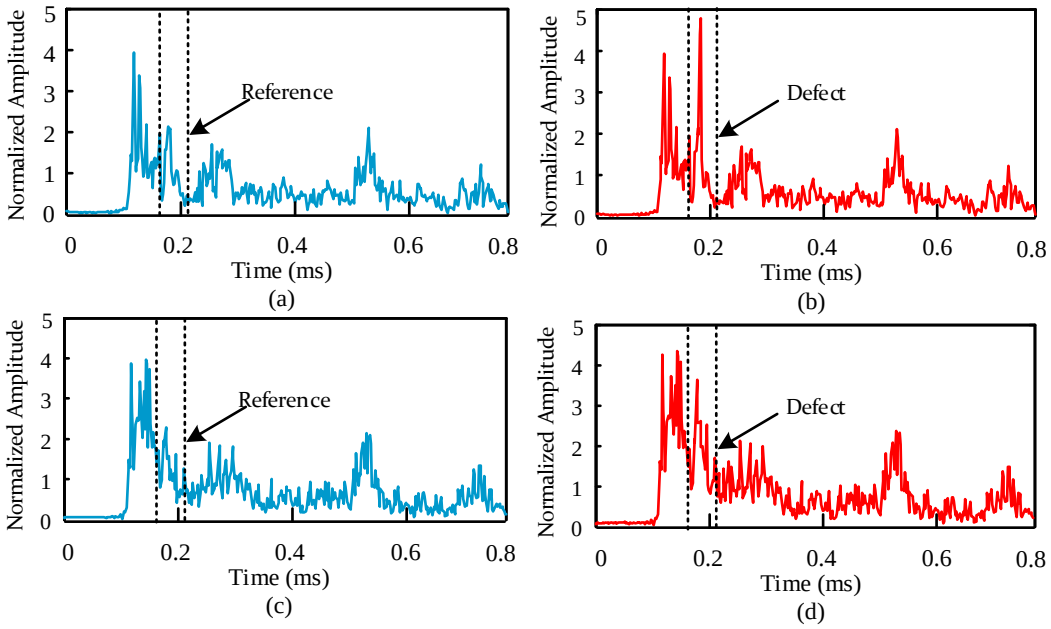


Figure. 9 The residual signal after the NARX process from: (a) a signal from $\mathbf{s}^{(2013)}$, (b) a signal from $\mathbf{s}^{(syn,1)}$ comprised of the signal in (a) and a synthetic defect at (1.1 mm, 0.3 mm) and with β = -30dB, (c) a signal from $\mathbf{s}^{(2020a)}$ and (d) a signal from $\mathbf{s}^{(2020b)}$ after the localised anomaly was added to the tank.

A common way to combine the residual signals, $e_{ij,k}(t)$, from different sensor pairs (i,j) for defect detection is to form an image using the ellipse imaging algorithm [29]. To better observe long-term trends, the average image from residual signals over a defined period is used, hence the intensity of the average image at position (x,y) is expressed as:

$$I(x,y) = \frac{1}{n_k}\sum_k \left| \sum_{i,j} a_{ij} e_{ij,k}^{(period)} \left(\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}+\sqrt{(x-x_j)^2+(y-y_j)^2}}{c_k} \right) \right|, \quad (15)$$

where $(x_{i,j}, y_{i,j})$ is the location of a sensor, a_{ij} is a weighting factor, n_k is the number of datasets in the period considered, and c_k is the mean estimated group velocity of the guided wave for that dataset based on measurements from each sensor pair (e.g. as shown in Fig. 3(b) for a single sensor pair). To balance the contributions from different sensor pairs, the value of a_{ij} is calculated from the maximum amplitude of the residual signals obtained in 2012 as $a_{ij} = 1/\max\left(\mathbf{e}_{ij}^{(2012)}\right)$. Figs. 10(a-g) compare the resultant average image of residual signals in various years. From Figs. 10(a-g), it can be seen that a response at (0.73 m, 0.46 m) appears gradually increasing in amplitude from 2012 to 2020. Physical inspection of the tank shows localised corrosion around this region. When the localised anomaly is added in 2020 a new localised indication is visible with globally highest amplitude at its location (1.1m,0.3m) in Fig. 10(g). This is most clearly seen when comparing the 2020 images in the immediate vicinity of the anomaly before and after its addition, as shown in Fig. 10(h). The RMS of the entire residual images in Figs. 10(f) and (g) are in a similar level (2.252 and 2.305), while the peak value within the 200 mm square region around the location of the anomaly shown in Fig. 10(h) increases abruptly from 3.477 to 5.774 when the anomaly is added.

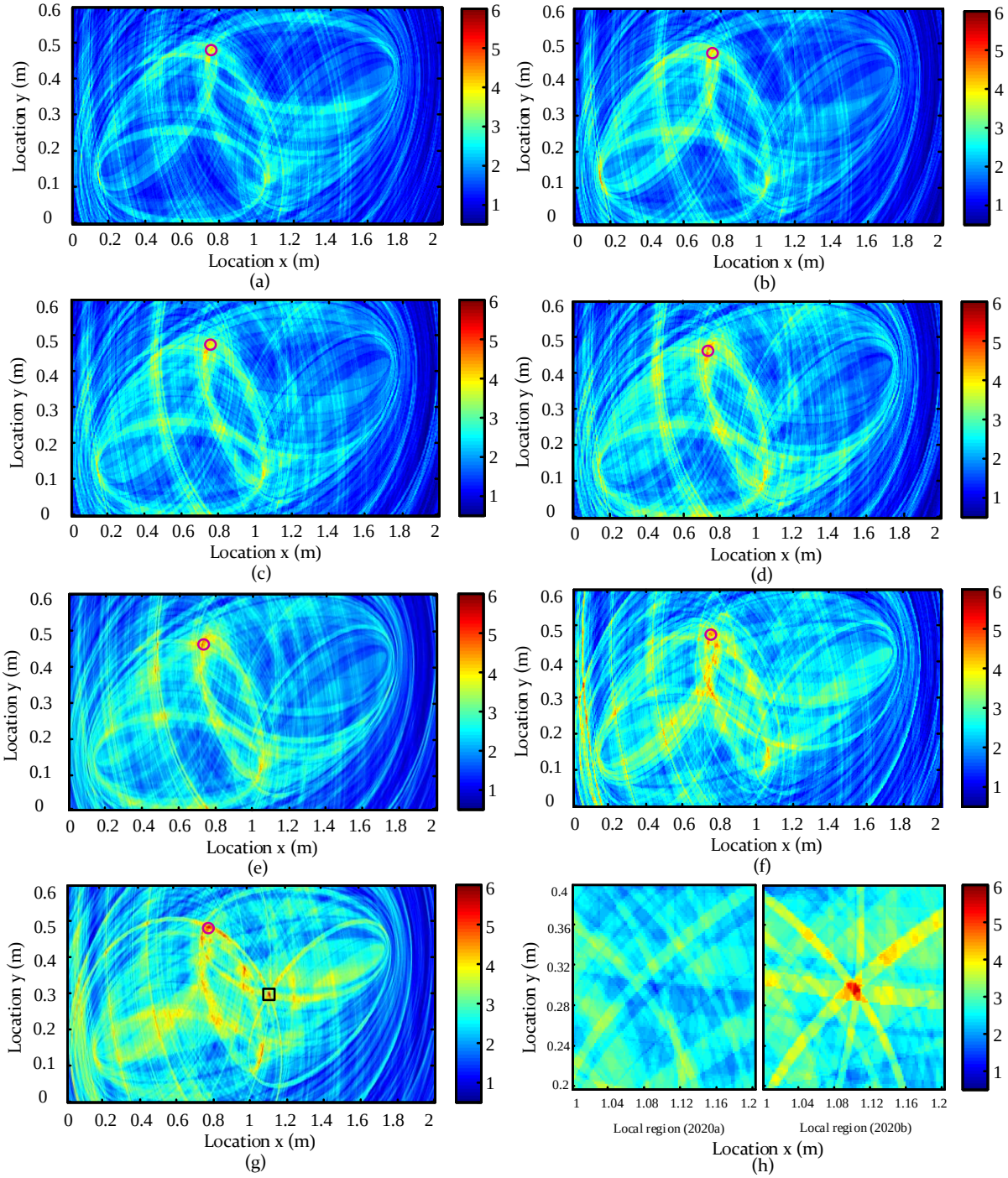


Figure. 10 Images formed using the residual signals after the NARX process from (a) $\mathbf{e}^{(2012)}$, (b) $\mathbf{e}^{(2013)}$, (c) $\mathbf{e}^{(2014)}$, (d) $\mathbf{e}^{(2015)}$, (e) $\mathbf{e}^{(2016)}$, (f) $\mathbf{e}^{(2020a)}$, (g) $\mathbf{e}^{(2020b)}$, and (h) subfigures of local region around the anomaly in (f-g). Empty purple circles and empty black squares correspond to locations of corrosion and anomaly, respectively.

Figs. 11(a-c) shows images generated using synthetic defect signals, $\mathbf{e}^{(syn,2)}$, (generated from 2020*a* data) with various β for a defect located at the same location as the physical localised anomaly. The image response at the defect location for each value of β is extracted and plotted in Fig. 11(d), which also

shows the response from the physical anomaly. From this it can be deduced that the physical anomaly corresponds to a synthetic defect of around β = -30 dB. As important as the peak amplitude is that the other image features close to the defect location (eg in Fig. 11(b)) agree well with the response seen in Fig. 10(g) and (h). Moreover, although piezoelectric sensors experience drift over years of operation as shown in Fig. 3, the proposed method is shown to be able to suppress the concept drift over time and detect the anomaly under this scenario. It is believed that the effects of network evolution are smaller than those caused by the environmental effects, and the training is to a certain extent accounting for both.

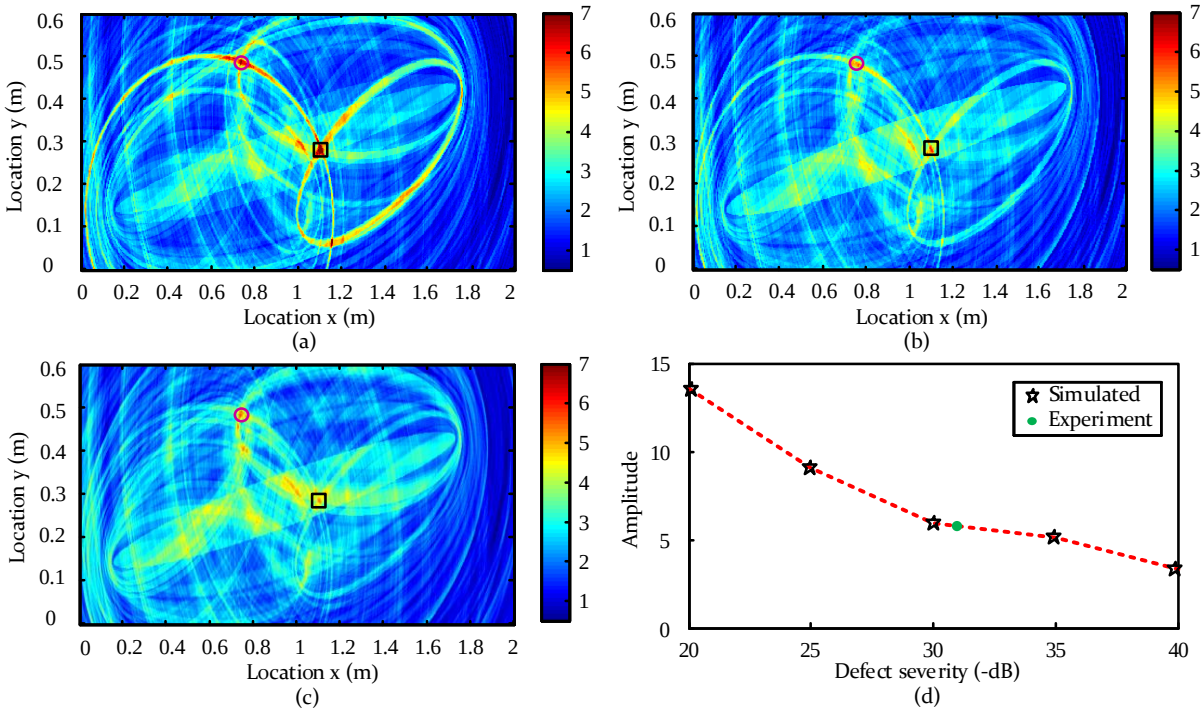


Figure. 11 An image formed using the residual signals after the NARX process on $\mathbf{e}^{(syn,1)}$ for the defect at the real defect location and with β = (a) -25 dB, (b) -30 dB, and (c)-35 dB. (d) defect amplitudes in the residual images.

4 CONCLUSION

The feasibility of using NARX networks to train a guided wave structural health monitoring system has been investigated. The procedure to design the NARX network, choose key parameters and test the network have been illustrated in detail. On the basis of Takens' theorem, the time delay and embedding dimension of the input layer are determined with mutual information and Cao's method. The required

size of training set is determined by increasing the number of training signals until no further significant improvement in performance is obtained. It is important to note that a larger than optimum training set be used in practice with no detriment to ultimate performance. The NARX method was applied to train networks using data obtained from different sensor pairs in a sparse array. These trained networks' detection performance were tested with both synthetic defect data (simulated defect responses superposed onto experimentally-measured data) and experimentally-measured data from a localised anomaly (added mass) introduced to the test structure.

The detection performance of the NARX was firstly compared using synthetic defect data with the standard Optimal Baseline Subtraction (OBS) technique, which uses the same set of training data as a baseline dataset. ROC curves are obtained that show that the NARX method has significantly better detection performance than OBS. A small set of 20 baselines was sufficient to account for the non-uniform changes in the EOCs the test structure experienced in real use. In tests using experimentally measured datasets acquired over several years, images were formed using the residual signals output by the NARX method from a sparse array of sensors. Images showed no obvious degradation in their baseline EOC suppression and did show a previously unobserved gradually increasing localized response that is believed to be due to localized corrosion. When the mass was added to the structure a new indication is seen at the correct location.

The use of the NARX method offers a powerful approach to the problem of GWSHM, sidestepping several of the key problems limiting machine learning applications to the field in general. For the purpose of modelling time series responses under specific known conditions, a small training set is sufficient to produce excellent EOC suppression and there is no need for very large quantities of labelled data to train models. Indeed, the use of the anomaly detector approach outlined here means that no specific knowledge of damage type is required to develop the network used.

It should be stressed that the NARX networks were trained using datasets captured in 2012 and tested on datasets captured from 2013 to 2020 with no significant degradation in performance. When similar measurements were made using OBS processed data previously the performance had degraded to the point to be useless by this stage. This shows the remarkable robustness of the NARX method used here proving stable with varying environmental conditions and the aging of the installed sensor network.

It should also be noted that the performance of the NARX method outside the range of EOCs experienced during trained has not been explicitly investigated. As the training samples were obtained over a complete annual cycle, it is to be expected that the EOCs seen in subsequent years would be within the same range. Whether or not the NARX method can deal with EOCs substantially outside the training range is an avenue for future investigation.

ACKNOWLEDGEMENTS

The work reported is part of a pilot project (Grant number 100374) funded by Lloyd's Register Foundation and the Alan Turing Institute Data-Centric Engineering Programme. Kangwei Wang was supported by the China Scholarship Council (Grant number 201906120066). Anthony J. Croxford was supported by an EPSRC fellowship (Grant number EP/M022528/1).

REFERENCES

1. J. L. Rose, M. J. Avioli, P. Mudge and R. Sanderson, "Guided wave inspection potential of defects in rail," NDT & E International 37, 153-161 (2004).

2. C. A. Chua and P. Cawley and P. B. Nagy, " Scattering of the Fundamental Shear Guided Wave from a Surface-Breaking Crack in Plate-Like Structures," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency control 66 (12), 1887-1897 (2019).

3. M. Mitra and S. Gopalakrishnan, "Guided wave based structural health monitoring: A review," Smart Materials and Structures 25 (5), 1-27 (2016).

4. E. Leinov, M. J. S. Lowe and P. Cawley, "Investigation of guided wave propagation in pipes fully and partially embedded in concrete," The Journal of the Acoustical Society of America 140 (6), 4528-4539 (2016).

5. M. Castaings and B. Hosten, "Ultrasonic guided waves for health monitoring of high-pressure composite tanks," NDT & E International 41, 648-655 (2008).

6. P. W. Loveday, C. S. Long and D. A. Ramatlo, "Ultrasonic guided wave monitoring of an operational rail track," Structural Health Monitoring-An International Journal, 19 (6), 1666-1684 (2020).

7. P. D. Wilcox, A. J. Croxford, N. Budyn, et al., "Fusion of multi-view ultrasonic data for increased detection performance in non-destructive evaluation, " Proc. R. Soc. A 476 (20200086), 1-26 (2020).

8. G. Konstantinidis, P. D. Wilcox and B. W. Drinkwater, "An investigation into the temperature stability of a guided wave structural health monitoring system using permanently attached sensors, " IEEE Sensors Journal 7 (5), 905-912 (2007).

9. A. J. Croxford, J. Moll, P. D. Wilcox and J. E. Michaels, "Efficient temperature compensation strategies for guided wave structural health monitoring, " Ultrasonics 50, 517-528 (2010).

10. C. A. Chua, and P. Cawley, " Crack growth monitoring using fundamental shear horizontal guided waves, " Structural Health Monitoring-An International Journal 19 (5), 1311-1322 (2020).

11. S. Mariani, S. Heinlein and P. Cawley, "Compensation for temperature-dependent phase and velocity of guided wave signals in baseline subtraction for structural health monitoring," Structural Health Monitoring-An International Journal 19 (1), 26-47 (2020).

12. S. Mariani, S. Heinlein and P. Cawley, " Location specific temperature compensation of guided wave signals in structural health monitoring," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency control 67, 146-157 (2020).

13. A. C. S. Douglass and J. B. Harley, " Model-based statistical guided wave damage detection for an aluminium plate," Structural Health Monitoring-An International Journal 19 (6), 1937-1950 (2020).

14. Y. Ying, J. H. Garrett Jr., I. J. Oppenheim and et al., "Towards Data-Driven Structural Health Monitoring: Application of Machine Learning and Signal Processing to Damage Detection, " Journal of Computing in Civil Engineering 27 (6), 667-680 (2013).

15. H. Liu and Y. Zhang, "Deep learning based crack damage detection technique for thin plate structures using guided lamb wave signals, " Smart Materials and Structures 29, 1-14 (2020).

16. D. F. Hesser, G. K. Kocur and B. Markert, "Active source localization in wave guides based on machine learning, " Ultrasonics 106, 1-11 (2020).

17. M. A. Vega and M. D. Todd, "A variational Bayesian neural network for structural health monitoring and cost-informed decision-making in miter gates, " Structural Health Monitoring-An International Journal, Ahead of Print 14 February 2020, DOI: 10.1177/1475921720904543.

18. X. Zhang, Q. Hao, K. Wang and et al., "An investigation on acoustic emission detection of rail crack in actual application by chaos theory with improved feature detection method, " Journal of Sound and Vibration 436, 165-182 (2018).

19. L. Yan, A. Elgamal and G. W. Cottrell, "Substructure vibration NARX approach for statistical damage inference, " Journal of Engineering Mechanics 139 (6), 737-747 (2013).

20. L. G. B. Ruiz, M. P. Cuéllar, M. D. Calvo-Flores and et al., "An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings, " Energies 9 (9), 684 (2016).

21. L. Cao, " Practical method for determining the minimum embedding dimension of a scalar time series, " Physica D 110, 43-50 (1997).

22. M. Kennel, R. Brown and H. D. Abarbanel " Determining the embedding dimension for phase-space reconstruction using a geometrical construction, " Physical Review A 45 (6), 3403-3411 (1992).

23. F. Takens, "Detecting strange attractors in turbulence, " Lecture Notes in Mathematics 898 (1), 366-381 (1981).

24. J. M. P. Menezes Jr. and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation, " Neurocomputing 71, 3335-3343 (2008).

25. H. T. Pham, V. T. Tran and B. Yang, "A hybrid of nonlinear autoregressive model with exogenous input and autoregressive moving average model for long-term machine state forecasting, " Expert Systems with Applications 37 (4), 3310-3317 (2010).

26. Y. Gao, S. Liu, F. Li and Z. Liu, "Fault detection and diagnosis method for cooling dehumidifier based on LS-SVM NARX model, " International Journal of Refrigeration 61, 69-81 (2016).

27. C. A. Perez-Ramirez, J. P. Amezquita-Sanchez, M. Valtierrz-Rodriguez and et al., "Recurrent neural network model with Bayesian training and mutual information for response prediction of large buildings, " Engineering Structures 178, 603-615 (2019).

28. R. J. Frank, N. Davey and S. P. Hunt, "Time Series Prediction and Neural Networks, " Journal of Intelligent and Robotic Systems 31, 91-103 (2001).

29. O. Putkis and A. J. Croxford, "Continuous baseline growth and monitoring for guided wave SHM, " Smart Materials and Structures 22 (055029), 1-11 (2013).

30. J. E. Michaels, S. J. Lee, A. J. Croxford, and P. D. Wilcox, “Chirp excitation of ultrasonic guided waves,” Ultrasonics 53 (1), 265–270 (2013).

31. F. Hardesty, Metals Handbook. Properties and Selection: Iron, Steels, and High-Performance Alloys. ASM International, 1990.

32. D. Keles, J. Scelle, F. Paraschiv and W. Fichtner "Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks, " Applied Energy 162, 218-230 (2016).

33. G. Marcjasz, B. Uniejewski and R. Weron, "On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks," International Journal of Forecasting 35, 1520-1532 (2019).

34. K. Wang, X. Zhang, Q. Hao and et al., " Application of improved least-square generative adversarial networks for rail crack detection by AE technique, " Neurocomputing 332, 236-248 (2019).

35. L. W. Schmerr, Fundamentals of ultrasonic nondestructive evaluation - a modeling approach. New York: Prenum press, 1998.

36. C. Holmes, B. W. Drinkwater and P. D. Wilcox, "Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation," NDT & E International 38, 701-711 (2005).